The acquiescence effect in responding to a questionnaire

Der Zustimmungseffekt bei der Beantwortung von Fragebögen

Abstract

Objective: Since the acquiescence effect can distort assessments, it is important to test techniques to quantify this effect.
Methods: The tendency of acquiescence is tested by means of a questionnaire. 2037 representatively selected subjects filled in the Multidimensional Fatigue Inventory (MFI-20), which consists of five subscales with positively and negatively formulated items. For each subject and for each subscale an acquiescence score was calculated based on the simple sum of the answers to the items of both orientations.
Results: Extreme acquiescence scores were rare (about 0.5%). All correlations between the acquiescence scores of the subscales were positive with mean values of 0.24, which indicates a certain degree of individual consistency in the acquiescence behavior. In the exploratory and confirmatory factorial analyses the polarity of the items was at least as meaningful as the contents of the subscales. Persons with high degrees of acquiescence were principally older and more depressed. By means of Rasch scaling procedures differences in threshold parameters for the response categories between positively and negatively oriented items were found.
Conclusion: The acquiescence scores derived by simple addition of item values proved to be well suited to clarify amount and conditions of the acquiescence effect.

Keywords: acquiescence, yes-set, judgement effect, judgement error, response set

Andreas Hinz[1]
Dominik Michalski[2]
Reinhold Schwarz[3]
Philipp Yorck Herzberg[1]

1 Department of Medical Psychology and Medical Sociology, University of Leipzig, Leipzig, Germany

2 Department of Neurology, University of Leipzig, Leipzig, Germany

3 Department of Social Medicine, University of Leipzig, Leipzig, Germany

Zusammenfassung

Zielsetzung: Der Zustimmungseffekt (Akquieszenzeffekt) kann Fragebogenergebnisse verzerren. Es soll eine Methode zu dessen Quantifizierung erarbeitet und erprobt werden.
Methodik: Die Zustimmungstendenz (Akquieszenz) wird anhand eines Fragebogens analysiert. 2037 bevölkerungsrepräsentativ ausgewählte Personen füllten das Multidimensional Fatigue Inventory MFI-20 aus, welches aus 5 Skalen mit jeweils positiv und negativ gepolten Items besteht. Durch einfache Addition der Antworten ohne Berücksichtigung der Polung der Items wurde für jede Person und jede Skala ein Zustimmungswert berechnet.
Ergebnisse: Extreme Zustimmungs- oder Ablehnungstendenzen sind selten (rund 0,5%). Die Korrelationen zwischen den Zustimmungswerten der einzelnen Skalen sind sämtlich positiv und liegen im Mittel bei 0,24, was auf eine gewisse individuelle Konsistenz des Zustimmungsverhaltens verweist. In der explorativen und konfirmatorischen Faktorenanalyse hat die Polung der Items im vorliegenden Fragebogen ein mindestens ebenso hohes Gewicht wie die inhaltliche Zuordnung zu den Skalen. Personen mit hoher Zustimmungstendenz sind vorrangig ältere und depressivere Personen. Mittels Rasch-Skalierung wurden Unterschiede in den Schwellenparametern für die Kategoriengrenzen bei positiv und negativ gepolten Items aufgezeigt.
Fazit: Das relativ einfache Zustimmungsmaß erwies sich insgesamt als gut geeignet, Ausmaß und Bedingungen des Zustimmungseffekts zu analysieren.

Introduction

Acquiescence (yes-set) describes the general tendency of a person to provide affirmative answers to items of a questionnaire, regardless of the content of the items [1]. Obviously, this can cause distortions in the assessments. This effect can be measured when the questionnaire contains items which are located at opposite sites of bi-polar constructs. Here it may occur that items or sub-scales, which are assumed to be negatively correlated because of their opposite character, are nevertheless statistically independent from one another. This can be impressively observed in mood research. Costa and Mc-Crae [2] developed the concept "positive affectivity – negative affectivity" and found that positive affectivity is not the reverse of negative affectivity. Instead, both dimensions are orthogonally related in the mood space [3], [4]. One possible explanation for such a result is the acquiescence effect, which prevents a negative correlation between these dimensions.

Acquiescence effects can be avoided when items are presented in a bipolar way with verbal explanations at both ends of the scale. However, most questionnaires use unipolar items, and questionnaires with a balanced proportion of positively and negatively formulated items are rare. One possible reason for this could be the impact of the acquiescence effect itself. Given the existence of such an effect, it will cause a reduction of the internal consistency of the scale. When an inversely oriented item is added to a set of items sharing the same direction, Cronbach's alpha will not increase, or the increase will be lower compared to adding an item of the same direction. This may contribute to the tendency of questionnaires often omitting the inclusion of items with reverse orientations. When the validity of a questionnaire is assessed using the correlations with other questionnaires, a balanced proportion of negatively and positively oriented items can also weaken this criterion.

Calculating the exploratory factorial analyses with two factors for questionnaires comprising items with different orientations typically yields the following solution [5], [6]: The unrotated solution presents a first factor with high (absolute) loadings of all items with the signs of the items representing their polarity. The second factor is less pronounced with lower loadings and equal signs of the items. Considering the rotated factorial solution, each of the factors represents one subset of items with uniform polarity.

Structural equation modelling using latent method effects [7], [8] has been intensively used to test acquiescence effects. Billiet and McClendon [9] demanded that the effect should be verifiable at least in a subsample of persons agreeing to positively and negatively oriented items. Moreover, the effect should be generalisable across several domains as well as time. They chose two different

contents (prejudice against foreigners and mistrust in politics) and assessed them with positively and negatively formulated items. The best fit was found for a model which assumes different acquiescence effects for both domains, allowing positive correlations between these effects. The correlation between these latent variables was 0.44. Self-esteem is a further subject with several studies concerning acquiescence [7], [10]. The best fits were found for a model including a method factor for the negative items. With the help of a latent-class analysis [11], persons with inconsistent response patterns (agreeing to both positive and negative items) can also be identified. This has been shown for the depression scale CES-D (Center for Epidemiologic Studies Depression Scale) [12]. The results indicate that persons with reduced mental abilities sometimes do not understand negative formulations and, therefore, respond inconsistently [13]. Further reasons of inconsistent response patterns are ambiguity, high difficulty of the topics to be judged or limited motivation and knowledge of the respondents [14]. Altogether, there is no doubt that acquiescence effects do exist, but there are open questions concerning amount and conditions of these effects.

The objective of this paper is to show that these questions can be adequately addressed using a simple descriptive measure of acquiescence for each subject. By means of a questionnaire with a balanced proportion of positive and negative items in several subdimensions the following questions will be answered:

- What are the parameters for the distribution of the acquiescence effect?
- Are acquiescence effects, which are obtained for several scales, mutually correlated (generalisability)?
- Does the acquiescence effect depend on the score of the person on the dimension?
- Does the acquiescence effect depend on age and gender, and what is the relation with other psychological dimensions such as depression?

For reasons of comparability, the results will be compared with structural equation models and Rasch models.

Construction of a measure for acquiescence

As stated above, acquiescence effects can be appropriately examined, when questionnaires compress several scales, in which each scale has the same number of positive and negative items. This will be demonstrated by means of the Multidimensional Fatigue Questionnaire MFI-20 [15]. The MFI-20 contains items with a five-point answer format. The end categories are verbally labelled (yes, this is true ... no, this is not true). Each of the five scales contains two positively oriented items with respect to fatigue (e.g., "I feel tired") and two negatively oriented items (e.g., "I am not tired at all", or "I feel very active").

The acquiescence score is calculated as follows: For each scale the sum over all items is calculated, without prior inversion of the negatively oriented items. In order to obtain a better comparability with questionnaires of other answer patterns and item numbers, the sum scores are linearly transformed, so that persons with maximal affirmation to all (positive and negative) items are assigned a score of 100 and persons with maximal rejection a score of 0. Acquiescence refers to a bipolar dimension here, ranging from extreme rejection (0) to extreme affirmation (100). A person has a medium acquiescence effect (50) when affirmation and rejection are equalized.

Acquiescence (measured over a scale) and the score on the scale (fatigue) are two distinct subjects. A person with neutral acquiescence (score 50) can have a low fatigue score, agreeing to all negative and refusing all positive items. Another person with an acquiescence score of 50 can be extremely tired. That is, acquiescence and "true" score of the scale are, at least theoretically, independent from each other. However, a person with an extreme score on the scale (e. g., extreme fatigue) must have a neutral acquiescence score (50), since an extreme fatigue score can only be reached when the person maximally agrees to all fatigue items and maximally rejects all opposite items.

Methods

Sample

In 1998 a survey was conducted with a sample which can be assumed to be representative of the German population living in private houses. 2037 persons were assessed (895 males, 1142 females). Ages ranged from 14 to 92 years, the mean age was 49.2 years (SD=17.3 years).

Instruments

The subjects were tested with the Multidimensional Fatigue Inventory MFI-20. This questionnaire measures several aspects of fatigue. It proved to be well applicable to assess the quality of life in cancer patients [16], [17]. The test consists of 20 items which belong to five scales: General Fatigue, Physical Fatigue, Reduced Motivation, Reduced Concentration, and Mental Fatigue. Each scale is represented by four items, two positively and two negatively oriented ones. High scores always designate a high degree of fatigue.

For the calculation of test scores, the items which indicate low fatigue have firstly to be inverted. After that, the mean score is calculated for each scale, resulting in a score between 1 and 5. Furthermore, a total score can be calculated, defined as the mean of the five scale values.

Beyond the MFI-20, several other questionnaires were used, e. g. the Hospital Anxiety and Depression Scale HADS [18] and the Sense of Coherence Scale [19].

Results

Magnitude of acquiescence

Table 1 shows the acquiescence scores for the five scales with their frequencies, calculated according to the procedure described above. The margin categories are given with higher accuracy (0.01%). The second column presents an example of a response pattern for the four items of the scale. Assigning answer 1 (not at all true) to all items gives rise to a minimal acquiescence score (0). Mean acquiescence can be found for response pattern 3333, but also for 2244 or 5151. As previously stated persons with maximal fatigue must have a neutral acquiescence score of 50.

Extreme respondents which only choose extremely low or high responses were rare and constituted less than 0.5% in all scales. When defining "pronounced rejection" as ranging from 6.25 to 25 (6.25 is the lowest possible score beyond 0) and "pronounced acquiescence" as ranging from 75 to 93.75, only 6% of the subjects belonged to one of these ranges. The majority of respondents was roughly balanced. About one third of the subjects scored with an acquiescence score of exactly 50. Table 1 also shows that the mean acquiescence value of the scales was about 50 and the corresponding standard deviation about 10. The mean acquiescence score across all five scales (20 items) was 47.3.

Pearson correlations among the acquiescence scores of the five scales ranged between 0.15 and .40, with an average of 0.24 (cf. Table 2).

Factorial structure

The correlations among the items are given in Table 3. Mean correlations can be calculated depending on the orientation of the items and the belonging to the scales. Mean correlation coefficients between items with equal orientation within one scale were 0.57; the corresponding coefficients for item pairs with different orientations within one scale were -.50. When items from different scales were considered, the mean correlation of items with equal polarity was 0.51, while inversely oriented items correlated with -0.44.

Furthermore, a Principal Component Analysis (PCA) with varimax rotation was performed. The highest eigenvalues were 10.2, 1.3, and 0.99. According to the eigenvalue >1 criterion, a two-factorial solution emerges. Therefore, we computed a two-factorial and a five-factorial solution (according to the number of subscales proposed by the test authors). Table 4 shows the loadings, the rotated solution for the five-factorial structure and both the unrotated and the rotated solution for the two-factorial solution. Loadings of ≥0.50 are underlined. Items are arranged according to scales and polarity to facilitate inspection. Negative items were not inverted.

Table 1: Distribution of acquiescence scores for the MFI-20 scales

Acquiescence score	Example	Scale 1 (%)	Scale 2 (%)	Scale 3 (%)	Scale 4 (%)	Scale 5 (%)
0.00	1111	0.05	0.10	0.25	0.30	0.10
6.25	1112	0.00	0.00	0.00	0.45	0.05
12.50	1122	0.15	0.10	0.25	1.33	0.15
18.75	1222	0.6	0.6	0.6	2.3	0.2
25.00	2222	1.9	3.6	2.7	5.9	1.2
31.25	2223	2.8	6.4	7.1	8.3	2.0
37.50	2233	10.6	16.0	16.6	16.8	7.4
43.75	2333	20.3	24.6	23.0	20.8	15.8
50.00	3333	34.5	36.2	32.9	28.1	38.4
56.25	3334	15.1	7.1	9.4	7.7	18.4
62.50	3344	7.4	2.9	3.7	3.8	8.5
68.75	3444	3.9	0.7	1.4	2.3	4.0
75.00	4444	1.9	1.1	1.4	1.4	2.5
81.25	4445	0.6	0.3	0.4	0.3	0.8
87.50	4455	0.20	0.15	0.10	0.15	0.30
93.75	4555	0.00	0.00	0.05	0.05	0.10
100.00	5555	0.00	0.05	0.15	0.15	0.15
Mean acquiescence		49.5	45.5	46.0	44.0	51.3
SD acquiescence		10.3	9.6	10.5	12.5	10.3

Table 2: Correlations between the acquiescence scores of the scales

Scale	Scale 1	Scale 2	Scale 3	Scale 4	Scale 5
Scale 1	-	.15	.25	.16	.15
Scale 2		-	.26	.28	.21
Scale 3			-	.40	.28
Scale 4				-	.28
Scale 5					-

N=2037; all coefficients significant with p<0.001

The results of the five-factorial solution do not confirm the theoretically assumed factorial structure. All scales are represented in several factors. One possible explanation for the obtained results lies in the polarity of the items. Factor 1 contains (among loadings of ≥0.50) only positively oriented items (from four hypothetical dimensions), and factor 2 only negatively oriented ones (from three dimensions).

The unrotated two-factorial solution presents high loadings in factor 1 according to the high first eigenvalue with signs alternating according to the item polarity. Factor 2 of the unrotated solution provides only unsystematic and weak contributions. In the rotated factorial solution factor 1 again presents the alternate form with highest loadings of the positively oriented items. However, factor 2 also contributes substantially with its signs always being inverted with respect to those of factor 1.

In a second step several CFA (confirmatory factorial analysis) models were tested using maximum likelihood estimation. Firstly, a basic model with five correlated factors according to the five scales of the MFI-20 was computed. The fit of the basic model was not sufficient (CFI=.894, TLI=.878, RMSEA [90% CI]=.089 [.086-.091]), indicating that adding method factors could enhance model fit. Therefore, a second CFA model consisting of five content factors and a common method factor was tested. In a first step the factor loadings of the method factor were freely estimated. In a second step the factor loadings of the method factor were constrained to be equivalent. Testing the differences between both models revealed a significant change in model fit for chi-square

Table 3: Correlations among the items

	Scale 1				Scale 2				Scale 3				Scale 4				Scale 5			
	p1	p2	n1	n2	p1	p2	n1	n2	p1	p2	n1	n2	p1	p2	n1	n2	p1	p2	n1	n2
s1p1	-	.50	-.48	-.51	.66	.66	-.64	-.57	.69	.54	-.47	-.52	.55	.46	-.38	-.37	.46	.47	-.35	-.35
s1p2		-	-.53	-.51	.55	.57	-.43	-.52	.52	.45	-.35	-.44	.45	.40	-.38	-.37	.49	.53	-.36	-.38
s1n1			-	.57	-.48	-.51	.49	.49	-.52	-.41	.41	.47	-.42	-.34	.41	.36	-.43	-.43	.37	.36
s1n2				-	-.54	-.58	.54	.58	-.54	-.46	.51	.63	-.44	-.41	.45	.47	-.46	-.49	.46	.44
s2p1					-	.68	-.61	-.60	.70	.63	-.52	-.56	.55	.57	-.45	-.40	.57	.52	-.40	-.38
s2p2						-	-.61	-.66	.66	.53	-.49	-.55	.53	.52	-.45	-.42	.49	.50	-.37	-.37
s2n1							-	.62	-.61	-.54	.54	.62	-.45	-.41	.45	.45	-.42	-.43	.41	.37
s2n2								-	-.58	-.50	.56	.61	-.45	-.43	.49	.53	-.44	-.48	.44	.41
s3p1									-	.62	-.51	-.57	.66	.55	-.43	-.42	.52	.51	-.35	-.36
s3p2										-	-.54	-.59	.51	.52	-.43	-.40	.57	.50	-.35	-.34
s3n1											-	.63	-.40	-.39	.53	.50	-.47	-.44	.46	.39
s3n2												-	-.44	-.43	.52	.58	-.46	-.49	.48	.45
s4p1													-	.54	-.37	-.33	.50	.48	-.31	-.31
s4p2														-	-.36	-.32	.45	.48	-.30	-.27
s4n1															-	.47	-.43	-.44	.45	.42
s4n2																-	-.41	-.42	.42	.44
s5p1																	-	.68	-.46	-.42
s5p2																		-	-.50	-.49
s5n1																			-	.42
s5n2																				-

N=2037; all correlations significant with p<0.001

s: Scale; p: positive item; n: negative item

Table 4: Factor loadings of the exploratory (left) and confirmatory (right) factorial analysis

Scale	Nr.	Direct.	5 Factors, rotated					2 Factors, unrotated		2 Factors, rotated		CFA (MTMM)		
			F1	F2	F3	F4	F5	F1	F2	F1	F2	**Content**	pos.	neg.
Scale 1	1	+	.62	-.30	-.46	.00	-.11	.75	.29	.76	-.27	.33	.81	-
	12	+	.34	-.04	-.66	.30	-.23	.68	.10	.58	-.37	.36	.62	-
	5	-	-.19	.23	.74	-.24	.05	-.67	.00	-.51	.43	-.45	-	.82
	16	-	-.23	.44	.57	-.21	.20	-.75	.13	-.48	.59	.03	-	.79
Scale 2	8	+	.66	-.32	-.37	.18	-.08	.81	.25	.78	-.33	.41	.86	-
	20	+	.56	-.32	-.52	.05	-.15	.79	.21	.74	-.35	.03	.71	-
	2	-	-.43	.57	.42	.02	.07	-.75	-.05	-.60	.45	-.25	-	.70
	14	-	-.36	.53	.47	-.04	.22	-.77	.06	-.55	.55	-.14	-	.64
Scale 3	3	+	.71	-.30	-.38	.08	-.11	.80	.32	.82	-.28	.36	.62	-
	6	+	.65	-.41	-.11	.28	-.02	.74	.16	.67	-.36	.38	.61	-
	10	-	-.30	.74	.12	-.24	.05	-.71	.23	-.39	.64	-.34	-	.62
	17	-	-.33	.69	.27	-.15	.21	-.78	.21	-.45	.66	-.03	-	.72
Scale 4	4	+	.72	-.11	-.21	.18	-.17	.68	.34	.74	-.19	.37	.81	-
	15	+	.73	-.17	-.06	.24	-.08	.64	.32	.69	-.18	.37	.76	-
	9	-	-.20	.57	.16	-.33	.21	-.65	.32	-.29	.66	-.18	-	.68
	18	-	-.17	.59	.12	-.10	.53	-.63	.41	-.22	.72	-.13	-	.77
Scale 5	7	+	.47	-.17	-.18	.66	-.19	.71	-.03	.52	-.48	.40	.58	-
	11	+	.42	-.14	-.25	.62	-.34	.72	-.09	.49	-.54	.43	.59	-
	13	-	-.01	.49	.26	-.63	.08	-.60	.44	-.17	.72	-.59	-	.52
	19	-	-.13	.22	.21	-.24	.81	-.57	.40	-.17	.70	-.52	-	.49
Explained Variance			21.5	17.4	14.3	9.4	7.1	50.9	6.5	32.1	25.3			

Table 5: Rasch Scaling: Left threshold parameter (T1) and widths of categories (Differences between adjacent threshold parameters for positive and negative items)

Scale	Positive items				Negative items				
	Item-No.	T1	Diff. S2-S1	Diff. S3-S2	Diff. S4-S3	Item-No.	T1	Diff. S2-S1	Diff. S3-S2	Diff. S4-S3
Scale 1	1	-1.14	0.79	0.76	0.09	5	-1.62	1.16	0.75	1.15
	12	-2.23	1.68	1.21	0.49	16	-1.39	1.06	0.81	0.39
Scale 2	8	-1.42	1.09	0.45	0.58	2	-.70	0.75	0.09	0.58
	20	-2.03	1.55	0.74	0.14	14	-.60	0.45	0.72	0.15
Scale 3	3	-1.74	1.30	0.70	0.58	10	-.64	0.19	1.05	0.37
	6	-1.71	1.20	0.68	0.90	17	-.83	0.68	0.65	0.67
Scale 4	4	-1.44	1.03	1.07	0.48	9	-.66	0.82	0.32	0.96
	15	-1.67	0.87	0.91	0.54	18	-.23	0.22	1.06	0.07
Scale 5	7	-1.14	1.25	0.99	0.48	13	-.79	0.71	0.46	0.73
	11	-1.25	1.26	0.98	0.34	19	-1.24	0.85	0.90	0.85
Mean		-1.58	1.21	0.85	0.46		-.87	0.69	0.68	0.60

($\Delta\chi^2$=356.94, df=19, p<.001), indicating the model with different loadings on the method factor to be more appropriate (CFI=.953, TLI=.939, RMSEA [90% CI]=.063 [.060-.066]). The next CFA model consisting of five content factors with five independent method factors failed to converge due to a non-positive covariance matrix. This also occurred in the case of five correlated method factors. The last CFA model consisted of five content factors and two method factors. The first method factor consisted of negative items, whereas the second method factor consisted of positive items. Two versions of this model can be specified: a model in which the covariance between the two method factors is constrained to be zero and a model in which the covariance between the two method factors is estimated freely. Testing the differences between both models revealed a significant change in model fit for chi-square ($\Delta\chi^2$=178.87, df=1, p<.001), indicating the model with correlated method factors to be more appropriate (CFI=.963, TLI=.952, RMSEA [90% CI]=.056 [.052–.059]). Factor loadings on the content factors and both method factors are depicted in Table 4.

Rasch scaling

The data set was Rasch analysed with the program WIN-MIRA [20]. First, negative items were inverted. Hence, high scores indicate high degrees of fatigue for all items. The ordinal (partial-credit-) Rasch model was used to detect threshold parameters for the response categories. Five categories result in four thresholds which assign the position of the response categories on the latent continuum. The results are given in Table 5. For reasons of clarity only the widths of the categories are presented. One can notice that the widths decrease with increasing categories when the items are positive (category 2 (1.21) > category 3 (0.85) > category 4 (0.46)). For all positive items category 2 is broader than category 4. Negative items do not show this effect, and the thresholds are generally higher than those of the positive items, especially for the lower thresholds.

Correlations between acquiescence effect and further variables

Table 6 presents correlations between acquiescence scores and several other variables such as age, gender, anxiety, depression, and sense of coherence. Old persons compared to young persons showed lower degrees of acquiescence. The effect varied among scales (between -0.04 and -.24) with a mean correlation of -0.18. The relationship between gender (dichotomous variable; males: 1; females: 2) and acquiescence was also tested with a correlation coefficient yielding no significant influence. The fatigue score itself (MFI-20) was weakly and negatively correlated with acquiescence: tired persons tended to reject items. While anxiety (HADS; Hospital Anxiety and Depression Scale) was independent from acquiescence, depression was negatively correlated with acquiescence. As for persons with high fatigue levels, depressed persons also tended to refuse items. Sense of coherence (SOC) was correlated with acquiescence in scale 3 (meaningfulness), with higher acquiescence scores for persons with high levels of SOC.

Table 6: Correlations between acquiescence and age, gender, fatigue (MFI), anxiety and depression (HADS) and sense of coherence (SOC)

Variables	Scale 1 acquiescence	Scale 2 acquiescence	Scale 3 acquiescence	Scale 4 acquiescence	Scale 5 acquiescence	Total acquiescence
Age	-.04	-.14**	-.08**	-.24**	-.05*	-.18**
Gender	-.01	-.04	-.02	-.01	.03	-.01
Fatigue (MFI total)	-.02	-.10**	-.07**	-.12**	-.04	-.11**
Anxiety (HADS)	-.04	.01	-.02	.01	.07**	.01
Depression (HADS)	-.11**	-.08**	-.11**	-.18**	-.01	-.16**
Comprehensibility (SOC)	.00	.03	.03	-.01	-.03	.00
Manageability (SOC)	.01	.06**	.06**	.07**	.01	.07**
Meaningfulness (SOC)	.07**	.09**	.12**	.19**	.05*	.17**
SOC total	.02	.06**	.07**	.09**	.01	.08**

*: $p < 0.05$;   **: $p < 0.01$

Discussion

The first objective of this study was to explore the amount of an acquiescence effect. Persons with extreme affirmation scores are rare. For the five scales tested only about 5 out of 2000 persons (0.2%) either maximally refused or agreed to all items (positive and negative). Only a single person made his/her choices consistently on the left or the right margin (see Table 1), whereas pronounced acquiescence or rejection (defined over the ranges ≥75 and ≤25) were found for 6% of the persons. Here we must acknowledge a certain degree of arbitrariness in the definition of these ranges. Persons with pronounced acquiescence behaviour should be handled with caution: When the questionnaire comprises positively and negatively oriented items, such persons with extreme acquiescence tendencies will produce an average score in the dimension to be assessed and hence not attract attention. However, when all items are oriented in the same direction, the assessment would overestimate or underestimate the score.

The mean acquiescence score of 47.3 indicates that questionnaires must not generally suffer from a general yes-set, because this score is even somewhat lower than the theoretical mean value of 50. Since we cannot assume that the positive and negative items have the same mean difficulty index, we cannot conclude that the value of 47.3 denotes a slight tendency to reject items.

Buse [21] found a mean acquiescence effect of 3.0 which is equivalent to a score of 52.8 in our standardisation. A possible reason for this somewhat higher degree of acquiescence may lie in the methodological approach, in which the negative items were constructed using the method of literal reversal [22]. Though this technique assures a maximum correspondence between positive and negative items with respect to the contents, the verbal conciseness or comprehensibility of the negative

items can be impaired. In such cases acquiescence can be due to differences in the verbal diction of the items. All correlations among the acquiescence scores of the subscales were positive, with a mean coefficient of 0.24. This verifies a certain degree of intraindividual consistency across topics. Similar correlations between acquiescence scores were also reported by Buse [21] with values between 0.22-0.32 and Vagt and Wend [23] with values between 0.30-0.46. The similarity of the subscales in our investigation could indicate an even higher association between the acquiescence scores, since acquiescence can individually vary depending on the topic [24]. Billiet and McClendon [9] tested acquiescence for two different topics and found an optimal fit for a model with different, but correlated acquiescence scores for both scales. The comparatively smaller coefficients obtained in our study (despite of the similarity in the content) probably result from the low number of items per dimension, since – according to the Spearman-Brown formula – the reliability of measurement (in this case: reliability of the acquiescence effect) increases with increasing number of items. The five-factorial structure of the MFI could not be confirmed. Neither exploratory nor confirmatory factorial analyses provided sufficient support for the model. Instead, the loading patterns were in line with the typical results [5]. Thus, one might ask why the factorial analyses were employed at all. Firstly, the role of polarity of the items in the factorial structure should be demonstrated, and secondly, it should be admitted that the five subscales did not clearly represent delimitable areas. This assertion was confirmed by CFA analysis. Furthermore, CFA analysis revealed that model fits could be obtained only when two method factors were included. The mean loading for the method factor for positive items was .71 and for negative items .69, whereas the mean loading for the content factors was .31. This indicates that the influence of polarity of the items was stronger for the determination of the factor structure of the MFI-20 than the item content.

The mean item correlations (Table 3), broken down by equality/inequality of subscales and item orientation, also indicate that both aspects, namely belonging to the subscales and item orientation, are of similar relevance. Equality of item orientation increases the correlation about 0.07 (compared with unequal orientation), and equality of subscales about 0.06. Due to the unclear subscale structure, it can be argued that the determination of acquiescence scores for these subscales was not adequate. However, even when the subscales can be circumscribed only incompletely, they are, nevertheless, useful for the calculation of acquiescence effects and provide information about the generalisability of the findings.

The results of Rasch scaling elicit differences in the use of the answer categories for positive and negative items. The lowest category threshold of the negative items was approximately halfway between the thresholds 1 and 2 of the positive items. Since the mean score of the items was about 2 (indicating a low degree of fatigue), the respondents expectedly display larger differentiation in the range of low fatigue, which is reflected by broader categories. It is surprising though, that this effect does not appear in the use of negative items. This indicates, that negative formulations can have an impact not only on acquiescence, but also on the use of the latent response categories and is supported by the finding that the mean correlations between negative items (r=0.48; cf. Table 3) are lower than those between positive items (r=0.56).

Persons with high degrees of acquiescence (cf. Table 6) tend to be young, non-depressed people with a low degree of fatigue and a high level of meaningfulness (sense of coherence). Gender and anxiety, on the other hand, have no influence on acquiescence behaviour.

The general advantage of the presented novel measure for acquiescence lies in its simplicity of calculation and usefulness in clarifying several essential questions of acquiescence, especially the relationship between acquiescence effects among several scales (generalisability), as well as relationships to the underlying construct (in this case: fatigue) and to other variables such as gender or depression. Furthermore, the assessment of the test-retest-reliability of this measure is also unproblematic in repeated examinations. However, in contrast to Rasch analyses, the measure is not sensible for different utilisations of the latent categories and compared to Structural Equation Modeling (SEM) no model testing is possible. Yet, the measure helps to raise the awareness of some open questions and problems which are not in the focus of SEM users.

Acquiescence is not the only response bias. Beyond a neutral acquiescence, a consistent respondent should present a low variance of his answers (after inverting negative items). This variance does not directly follow from the acquiescence: A person with a mean acquiescence score (50) can either have a minimal variance (3333) or a high variation (1155) in his responses. Nevertheless, variance and acquiescence are not independent from each other, as persons with extreme acquiescence scores always demonstrate inconsistent answer patterns. The construction of a variance-based score for judgement consistency and the clarification of the relationship between acquiescence and this variance-based measure of consistency remains as a task for the future. Furthermore, one has to distinguish precisely between negatively directed and negatively phrased items. In the MFI-20, four of the 10 negative items are constructed with the use of verbal negation (not, no, few, low). These four items are not sufficient for a systematic comparison with the remaining six items without negative phrases. A questionnaire with a systematic variation of items with and without negative wording within one direction of the scale would contribute to a better understanding of this effect.

The data set reported here is restricted to the topic of fatigue. The generalisability of the effects to other topics cannot be evaluated here. For this purpose other questionnaires with systematic balances of positively directed and negatively directed items are mandatory. Moreover, the assessment of the temporal stability of the acquiescence effect is a desirable task for future research.

Notes

Conflicts of interest

None declared.

References

1. Messick SJ. The psychology of acquiescence: An interpretation of research evidence. In: Berg IA, editor. Response set in personality assessment. Chicago: Aldine; 1967. p. 115-45.

2. Costa PT Jr, McCrae RR. Influence of extraversion and neuroticism on subjective well-being: Happy and unhappy people. J Pers Soc Psychol. 1980;38:668-78.

3. Russell JA, Carroll JM. On the bipolarity of positive and negative affect. Psychol Bull. 1999;125:3-30.

4. Watson D, Clark LA. The measurement and mismeasurement of mood: Recurrent and emergent issues. J Pers Assess. 1997;86:267-96.

5. Sjöberg L, Perrson E. The measurement of mood. Scand J Psychol. 1979;20:1-18.

6. Matschinger H, Krebs D. Zum Problem der Abbildung eindimensional konzipierter Konstrukte bei entgegengesetzter Antwortpolung. ZA-Information. 1998;43:81-110.

7. Corwyn RF. The factor structure of global self esteem among adolescents and adults. J Res Pers. 2000;34:357-79.

8. Horan PM, DiStefano C, Motl RW. Wording effects in self-esteem scales: Methodological artifact or response style? Struct Equ Modeling. 2003;10:435-55.

9. Billiet JB, McClendon MJ. Modeling acquiescence in measurement models for two balanced sets of items. Struct Equ Modeling. 2000;7:608-28.

10. Tomas JM, Oliver A. Rosenberg´s self-esteem scale: two factors or method effects. Struct Equ Modeling. 1999;6:84-98.

11. Rost J. Testtheorie, Testkonstruktion. Bern: Huber; 1996.

12. Matschinger H, Schork A, Riedel-Heller SG, Angermeyer MC. Zur Anwendung der CES-D bei älteren Menschen: Dimensionsstruktur und Meßartefakte. Diagnostica. 2000;46:29-37.

13. von Collani G, Herzberg PY. Zur internen Struktur des globalen Selbstwertgefühls nach Rosenberg. Z Diff Diag Psychol. 2003;24:9-22.

14. Krosnick JA. Response strategies for coping with the cognitive demands of attitude measures in surveys. Appl Cogn Psychol. 1991;5:213-36.

15. Smets EMA, Garssen B, Bonke B, De Haes JCJM. The multidimensional fatigue inventory (MFI): Psychometric qualities of an instrument to assess fatigue. J Psychosom Res. 1995;39:315-25.

16. Smets EMA, Garssen B, Cull A, De Haes JC. Application of the multidimensional fatigue inventory (MFI-20) in cancer patients receiving radiotherapy. Br J Cancer. 1996;73:241-5.

17. Schneider RA. Reliability and validity of the Multidimensional Fatigue Inventory (MFI-20) and the Rhoten Fatigue Scale among rural cancer outpatients. Cancer Nurs. 1998;21:370-3.

18. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiat Scand. 1983;67:361-70.

19. Antonovsky A. The structure and properties of the Sense of Coherence Scale. Soc Sci Med. 1993;36:725-33.

20. von Davier M. (2001). http://www.winmira.von-davier.de/ . (10 June 2007).

21. Buse L. Kritik am Moderatoransatz in der Akquieszenz-Forschung. Psychologische Beiträge. 1980;22:119-27.

22. Jackson DN. Acquiescence response styles: Problems of identification and control. In: Berg IA, editor. Response set in personality assessment. Chicago: Aldine; 1967. p. 71-114.

23. Vagt G, Wendt W. Akqieszenz und die Validität von Fragebogenskalen. Psychologische Beiträge. 1978;20:428-39.

24. Ray JJ. Reviving the problem of acquiescent response bias. J Soc Psychol. 1983;121:81-96.

Corresponding author:

Dr. Andreas Hinz

Department of Medical Psychology and Medical Sociology, University of Leipzig, Philipp-Rosenthal-Str. 55, 04103 Leipzig, Germany, Tel.: +49-341-97 18820, Fax: +49-341-97 18809

andreas.hinz@medizin.uni-leipzig.de